ESA/STAT/AC.234/21
27 April 2011

UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION

Expert Group meeting on International economic and social classifications
New York, 18-20 May 2011

Classification Management Systems in Statistics New Zealand

Andrew Hancock, Statistics New Zealand

Classification Management Systems in Statistics New Zealand

Discussion paper presented to the United Nations Expert Group on International Economic and Social Classifications Meeting

18-20 May 2011
New York, United States of America

Andrew Hancock
Classifications and Standards, Statistics New Zealand
Private Bag 4741
Christchurch, New Zealand
info@stats.govt.nz

www.stats.govt.nz

Introduction

This paper discusses the Classification Management System currently used within Statistics New Zealand by looking at the background to the original development of such a system, and highlighting how the system currently works and related issues. The paper will also discuss the potential redevelopments to take place to upgrade how Statistics New Zealand manages its statistical classifications and delivers them to users.

Background

As part of the Information Technology (IT) redevelopment programme in 1996 to move Statistics New Zealand's IT capability from a mainframe to a client/server environment, a classification management system was developed. This system, called the Classifications and Related Standards system (CARS), is a Sybase/Centura application designed as a centralised repository to store all classifications, concordances, synonyms for coding, and SAS formats used within Statistics New Zealand, and also to act as a repository for classifications used across the New Zealand Official Statistical System (OSS).

Originally scoped as a complete metadata system to address the emerging metadata needs within Statistics New Zealand, the project was de-scoped to be a classification repository with links to statistical standards and the ability to produce coding indexes and customised classification reports. However some functionality, whilst identified as useful was not able to be included in the final application. This included a survey link, user notification system and multi-concordance functionality.

The CARS system was introduced into the Statistics New Zealand IT production environment in November 1996 and has remained the central repository for all classifications related material used within Statistics New Zealand. There have been no major changes to the CARS data model since 1996 and the system has been robust and reliable. Upgrades have been done over time to accommodate technology changes or needs, for example changes in Sybase versions, Y2K and Windows changes.

What is CARS?

CARS is an on-line computer package used when storing and working with classifications which:

- Provides centralised classification storage, maintenance and access facilities for all classification data used in the development and processing of surveys;
- Facilitates the use of standard classifications in all statistical data;
- Facilitates the comparison and analysis of data by storing concordances;
- Provides common ways to update and access classifications data.

CARS sits in a development, test and production server environment. Every staff member of Statistics New Zealand has access to the application either in browse mode or through password entry. Browse mode allows users to see and use all released and current classifications only. Password mode allows users to edit, maintain and update classification and concordance information in the application subject to certain security privileges.

CARS Administration

CARS is administered by the Classifications and Standards team (C&S), which is the classifications business unit of Statistics New Zealand. C&S take responsibility for the maintenance and usage of the CARS system, and take ownership of all New Zealand Standard and other Standard classifications used across the Official Statistical System. The benefits of having a centralised classifications unit supported by a central classification repository ensures that there is consistency across classifications and concordances, and that standardisation can be progressed across statistical inputs and outputs.

C&S staff administer CARS through the CARS Supervisor group, which is a group responsible for creating the framework of the classification within CARS, and administering all aspects of the CARS system including training.

Each classification is defined with a unique abbreviation, title and then attributes such as codes, descriptors, labels and edit values are assigned as appropriate by the CARS Supervisor Group. This ensures a level of consistency and also allows a rudimentary monitoring of the use of standard classifications within Statistics New Zealand.

All classifications are owned by an owner user group. This group can include a single person from an organisational business unit and is responsible for maintaining the classifications, versions and concordances that it owns. Membership of an owner user group enables staff to have write access to edit and update classifications and create new versions or concordances.

Types of Classifications

The CARS system currently stores over 4,500 classifications. These are classified as:

(a) international standards such as ISIC, ISCO, ISCED, CPC;
(b) Australian and New Zealand standards such as ANZSIC, ANZSCO, ANZSRC;
(c) New Zealand standards such as NZSCNPO, NZSCSA, NZSCC;
(d) standard classifications such as Language, Ethnicity, Iwi;
(e) independent classifications which are survey specific and for which no standard currently exists.

All classifications used within Statistics New Zealand go through a corporate approval process before being stored in CARS, with the exception of the international classifications which are stored for reference purposes only.

Standard statistical classifications are grouped into three categories:

(a) Australian and New Zealand Standards – these are harmonised classifications jointly produced by Statistics New Zealand and the Australian Bureau of Statistics. These may be based on an international classification modified to Australian and New Zealand requirements. They are used across the Official Statistical System, require significant external and internal input to concepts and structure, and must be approved for use by the New Zealand Government Statistician and the Australian Statistician.
(b) New Zealand Standards - These may be based on an international classification modified to New Zealand requirements. They are used across the Official Statistical System, require significant external and internal input to concepts and structure, and must be approved for use by the New Zealand Government Statistician.
(c) Standard classifications - These may be based on an international classification modified to New Zealand requirements, or reflect a common approach across a number of statistical

surveys. They may be used across the Official Statistical System, may have external input to concepts and structure and must be approved for use by the New Zealand Government Statistician.

Independent classifications are those that are specific to one survey or output, and for which there is no existing or available standard classification. As independent classifications are picked up and used across more than one statistical survey, they may progress to being a standard classification, then a New Zealand Standard, and possibly become an Australian and New Zealand Standard.

CARS Navigator Screen

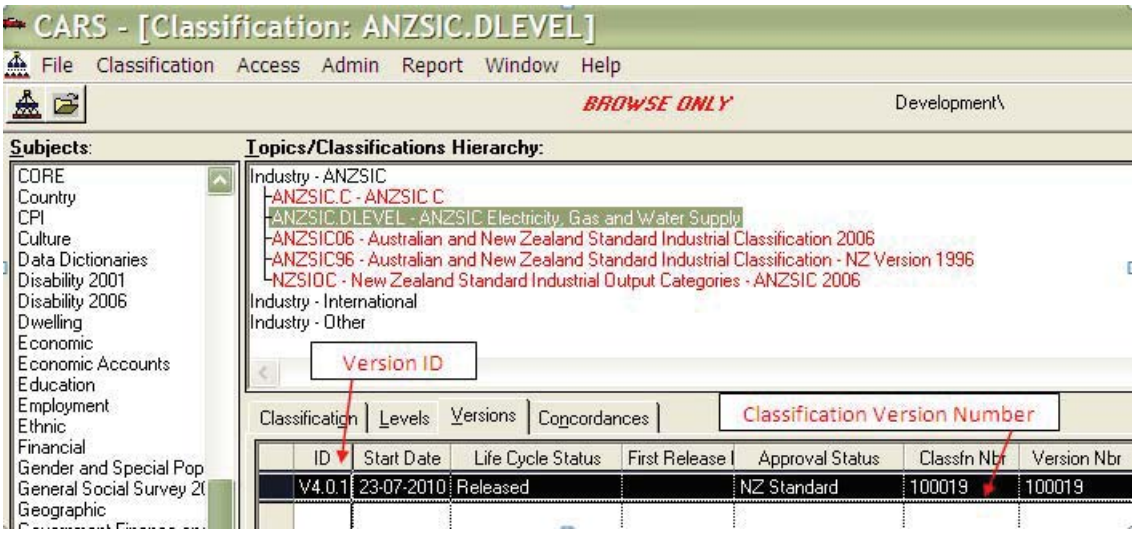
CARS is set up like a hierarchical classification system where users step down through subjects and topics to find a classification and its associated versions and concordances. Subjects are made up of statistical activities, broad subject matters or user defined subjects and within each of these are more detailed topics.

Each classification is given an unique abbreviation of up to 15 characters, a title (subject to naming conventions), a version identifier, a life cycle status (which indicates whether in development, released or obsolete), an approval status (which indicates whether standard or not), a release date and a system generated classification and version numbers (see Table 1).

Current released standard classification text is in the colour red – this was done to facilitate quick identification for users of the relevant classifications to be used. There is no mandatory enforcement for the use of standard classifications within Statistics New Zealand however subject matter areas are strongly advised to use them. Where they cannot, a business case for an exemption or use of another classification must be presented to the Standards Governance Board, which is the corporate approval body for classifications and standards.

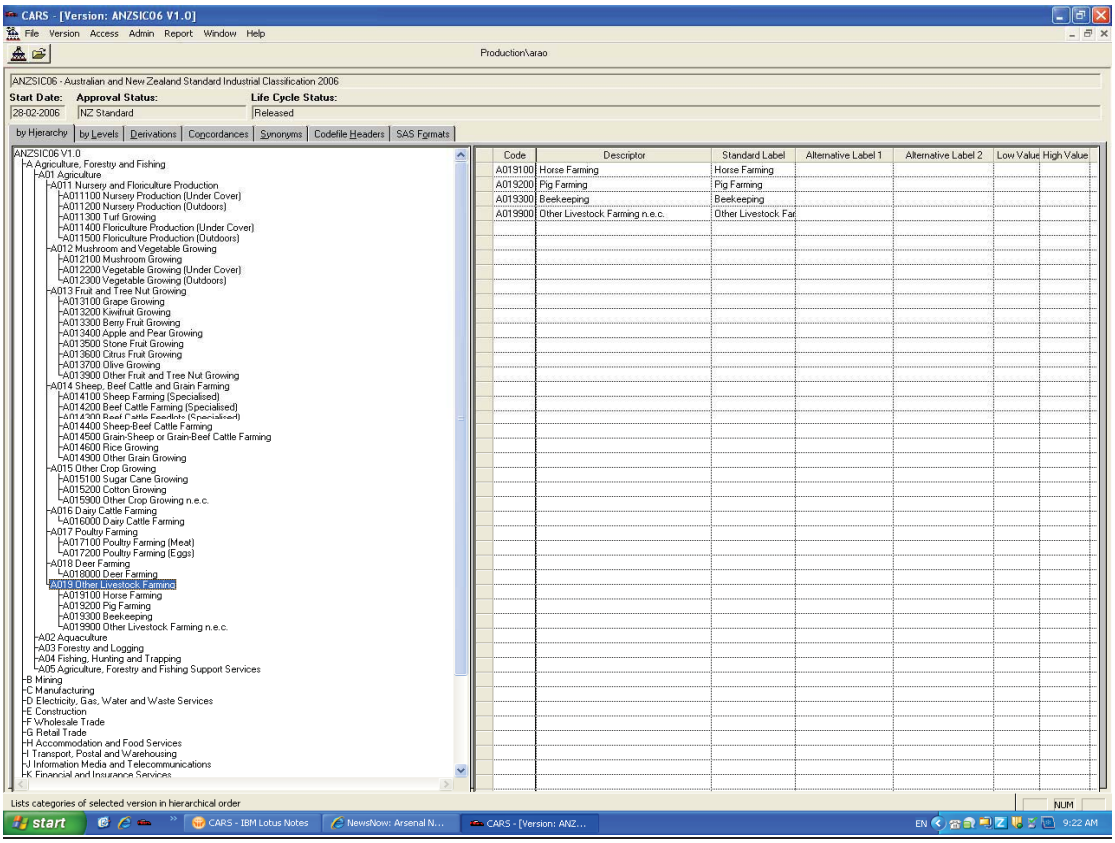
There are a number of protocols and conventions used within CARS. Classification titles adhere to rules specified in the Classifications Best Practice Guidelines document, which is the overarching document on how to produce statistical classifications. Survey specific classifications are given a three character prefix denoting the survey as part of their abbreviation. Classifications can be read restricted so that only selected staff can access or view them. This approach is sometimes utilised with administrative classifications that are used to supply data to Statistics New Zealand from external agencies.

Table 1: Existing CARS Navigator Screen



By double clicking anywhere on the highlighted line showing the relevant ID, Start Date etc, the user is then taken into a second screen (see Table 2 below) which enables viewing of the classification structure.

Table 2: Existing CARS Classification Hierarchy Screen



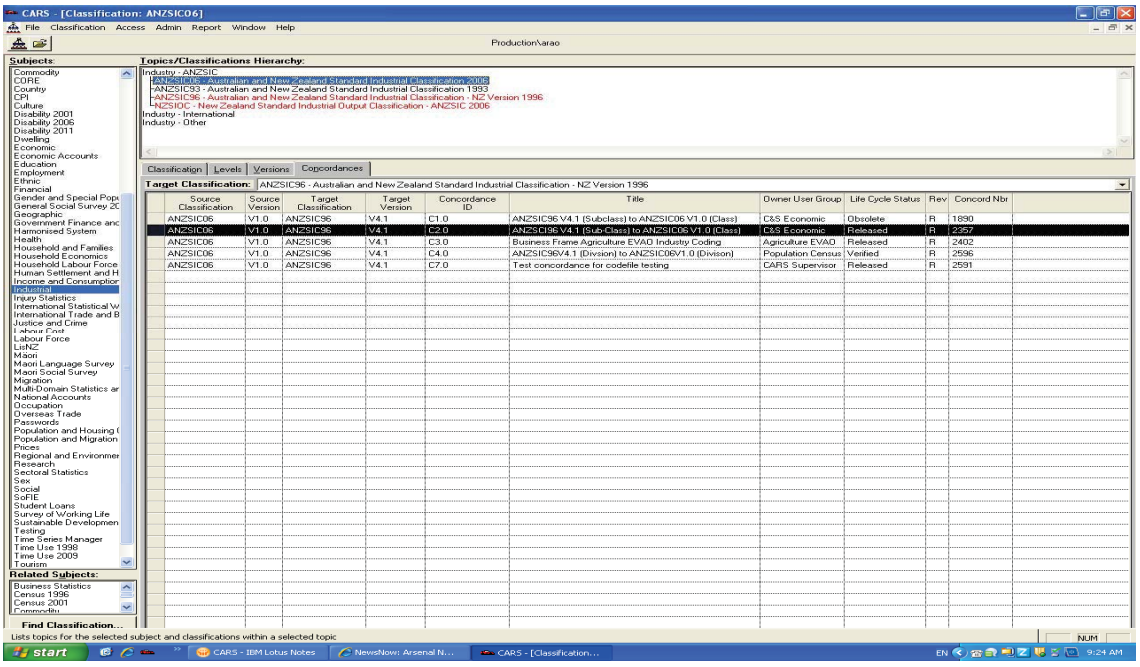
The default view is the top level of any hierarchical classification, or the full listing of a flat classification. For hierarchical classifications, the user can then double click on a code and descriptor, to view subsequent levels of the classification. The right hand part of the screen then shows additional information such as the standard labels or edits values that may be associated with each category. By selecting the properties of a category the user can then view any definition text or exclusion text relevant to the category.

Concordances

Concordances which map different classifications together or map different versions of the same classification are also stored in the CARS system. This facilitates the management of time-series and international reporting requirements. Concordances between international and New Zealand Standard classifications, and concordances between standard classifications are a mandatory requirement and these are usually developed and maintained by the Classifications and Standards team. Independent classifications may be concorded but this is not usually the case unless there is a time-series requirement for a specific survey.

Multi-concordances eg ISIC to SITC to BEC to CPC are not able to be stored in the current CARS system. This means that only one to one classifications or versions can be stored eg ISIC Rev 3 to ISIC Rev 4, or CPC V1.0 to CPC V2.0. The lack of multi-concordance functionality in CARS resulted in a Microsoft Access application being created that pulls concordances out of CARS and allows creation of multi-concordances.

Table 3: Existing CARS Concordance Screen



Coding Tools and Support

CARS also has an integrated coding tool – the Classification Coding System (CCS) which enables standardised coding to be utilised within survey processing. The synonyms are stored in CARS within a classification version and can be updated and coding indexes generated on demand. The CCS is available as a downloadable from the Statistics New Zealand website. Additionally CARS supports

the Classification Code Finder which is a web enabled coding application also available on the Statistics New Zealand website.

http://www.stats.govt.nz/tools_and_services/tools/ClassificationCodeFinder.aspx

The CARS system has a corporate approved policy requiring its use for all classifications information. There are a number of policies and protocols supporting its use and regular staff training programmes are run to enable staff across the organisation to use the application.

International Recognition

The original data model for CARS was supplied to Statistics Slovenia, which then formed the basis of their classification system (Klasje), and was also supplied to Statistics Austria. A customised CARS application was sold to the Central Statistical Office in Ireland, along with the Statistics New Zealand Business Frame and Balance of Payments system. The CARS data model and application was then converted into Microsoft SQL Server version which was sold to Statistics South Africa.

Future Developments for Classification Management in Statistics New Zealand

The existing CARS system was introduced to Statistics New Zealand in 1996 and has had no modifications or enhancements added since that time. Whilst CARS still performs as designed, and meets the primary needs for classification management, many IT systems that use CARS have been modified or business unit requirements have changed resulting in some need for work-arounds. This has most clearly been identified in the increasing production of output views of classifications to reflect user driven survey outputs which are not easily dealt with in the current CARS system. Another issue appearing is the increasing need to build multi-concordances between related classifications as opposed to classifications that are the same, for example, industry to harmonised system to broad economic categories, compared to industry to industry.

Some of the originally scoped aspects of CARS that were not included in the production system are now required, especially as the organisation works to develop a corporate metadata solution. The ability to easily link surveys to the classifications they use, notify users of changes as they happen, and to easily search across the database were unable to be included. A survey of users highlighted that these functions are essential components for any redevelopment.

In order to investigate requirements, a CARS 'proof of concept' (PoC) was created to demonstrate the potential benefits and efficiencies of upgrading the system, and demonstrate a platform for a full redevelopment of CARS. Some of the high level benefits were seen as:

- Facilitating use of standard classifications within administrative data sources across the Official Statistical System;
- Providing a centralised, visible and integrated solution for statistical infrastructure and metadata components;
- Enabling better relationships between statistical information and reporting on classification use;
- Increasing effectiveness and efficiency of classification, concordance and codefile management which would release resources to carry out more value-added tasks;
- Making classifications and standards easily accessible on the web and enabling standardisation across the Official Statistical System.

Key Aspects of the CARS 'proof of concept'

One of the primary concerns introducing a new system into an IT environment managing legacy system issues is to ensure that information created and stored in a new management system can be migrated back to the old system. This would enable existing infrastructure to continue to operate with classifications information until these systems were redeveloped.

Other changes proposed to the system are:

- Platform change from Sybase/Centura to Microsoft SQL Server 2008
- A classification view allowing recoding of classifications
- Improved navigation and search capability
- Multi concordance functionality
- Improved usage and notification management
- More linked internal systems

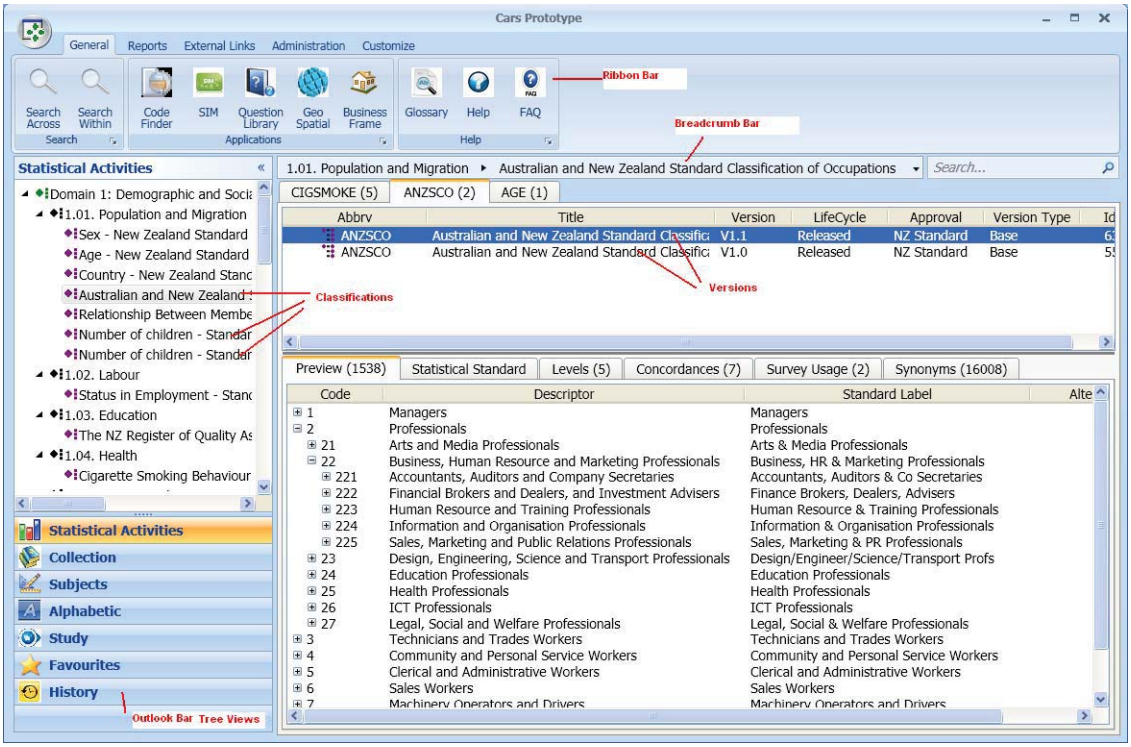
- Better support for storing statistical standards information
- Automatic web publishing of classifications, standards, and categories.

In 2010 a 'proof of concept' (PoC) version or prototype of a potential new CARS was developed and included some of the above features.

NB: The discussion that follows is based on the proof of concept system developed. Concepts, names of attributes and definitions presented are those used for the proof of concept exercise and should not be regarded as the final product. There are a significant number of metadata and classifications best practice issues to be addressed in a new CARS system should one be developed. It should also be noted that most of the existing functionality would be carried across so the following discussion focuses on the enhancements or changes proposed.

Navigator Screen

The PoC revamped the navigator screen to be more in keeping with current Microsoft products.

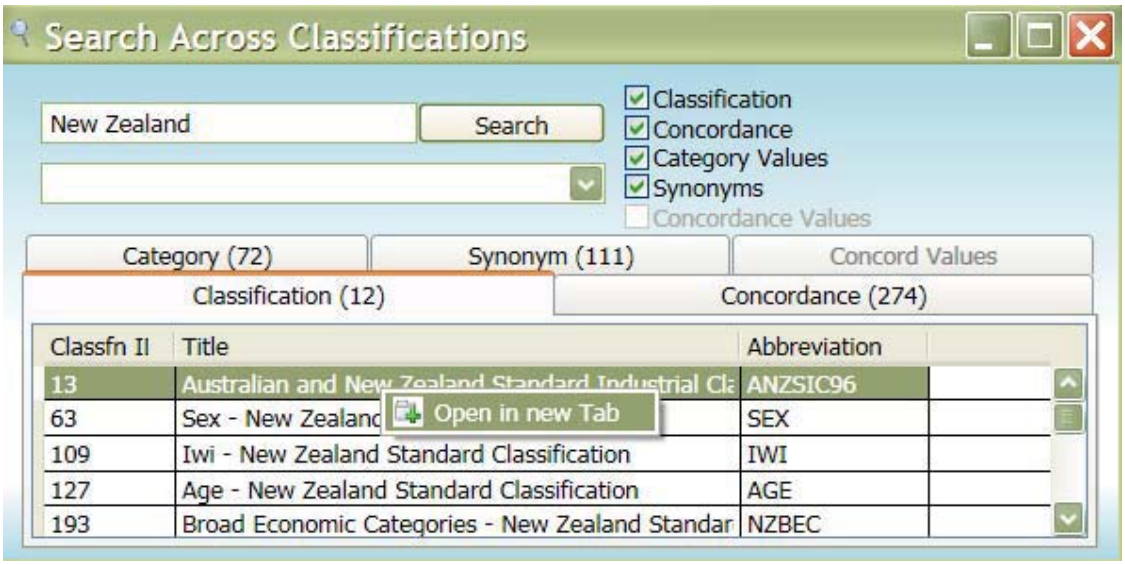


There are a number of ways to view classifications (as can been seen in the bottom left corner of the illustration above. The primary view is by statistical activities which uses the domains and activities prescribed in the United Nations Classification of Statistical Activities, which was adopted as a New Zealand Standard Classification in 2010. This view also allows SDMX compatibility. Other views are by statistical collection, subjects (as per the current CARS), alphabetically, or by user defined favourites or browsing history.

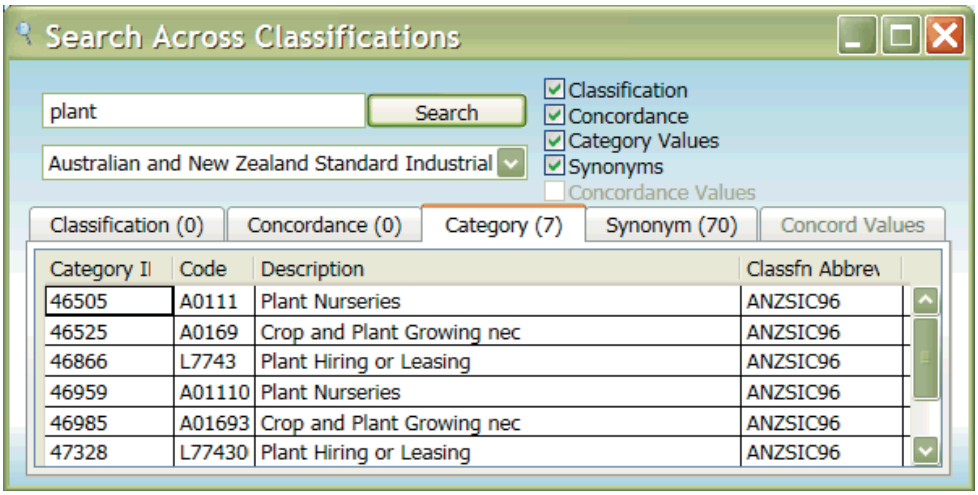
Available classifications are viewable in the left hand column and then displayed in the top of the navigator screen with a preview of the full classification at the bottom. Users are then able to view the supporting statistical standard, concordances, survey usage of the classification and synonyms (used for creating coding indexes).

Searching Across Classifications

Searching across classifications in the CARS PoC is very easy with a standard search engine capability included as shown in the following illustration.



The ability to search for a word or category within a classification is part of the existing CARS system but this facility has been enhanced to also allow searching across concordances and synonyms for coding. Previously users were required to extract concordances and or coding indexes into Microsoft Excel or Microsoft Access to be able to find specific words in these two important components.



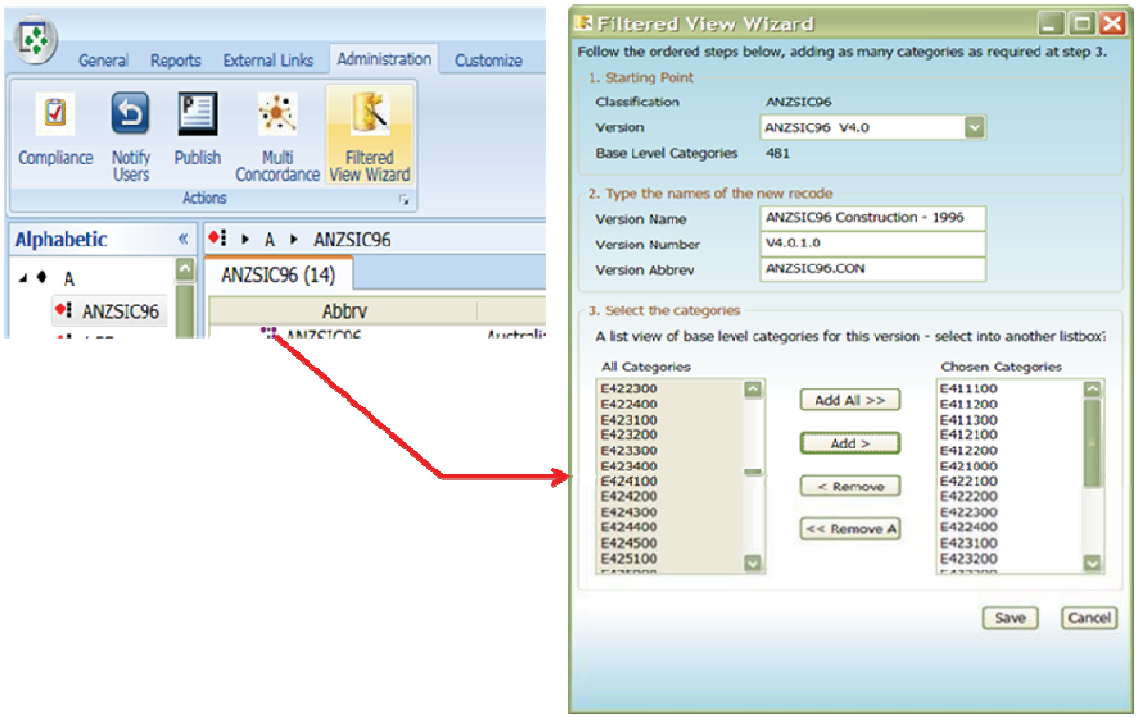
Filtered and Recoded Classifications

To address the increased demand for different types of output classifications, an option which is currently called 'Filtered and Recoded Classifications' has been introduced into the PoC. This gives users the flexibility to create different views of classifications, or create different aggregations of a classification.

- **Filtered Classification** is a portion of the original classification with no changes to the codes or descriptors.
- **Recoded Classification** is a Parent/Child relationship between codes is not related i.e. codes are not sequential.

A filtered view is not necessarily the equivalent of an alternative view.

From the Filtered View Wizard it is just a case of selecting the classification and then selecting the categories required for the recode from the 'All Categories' box and add them to the 'Chosen Categories' box and click save.



The resulting filtered views can then be stored in the CARS PoC and be accessible under the view of the parent classification. The same process can apply to creating recodes.

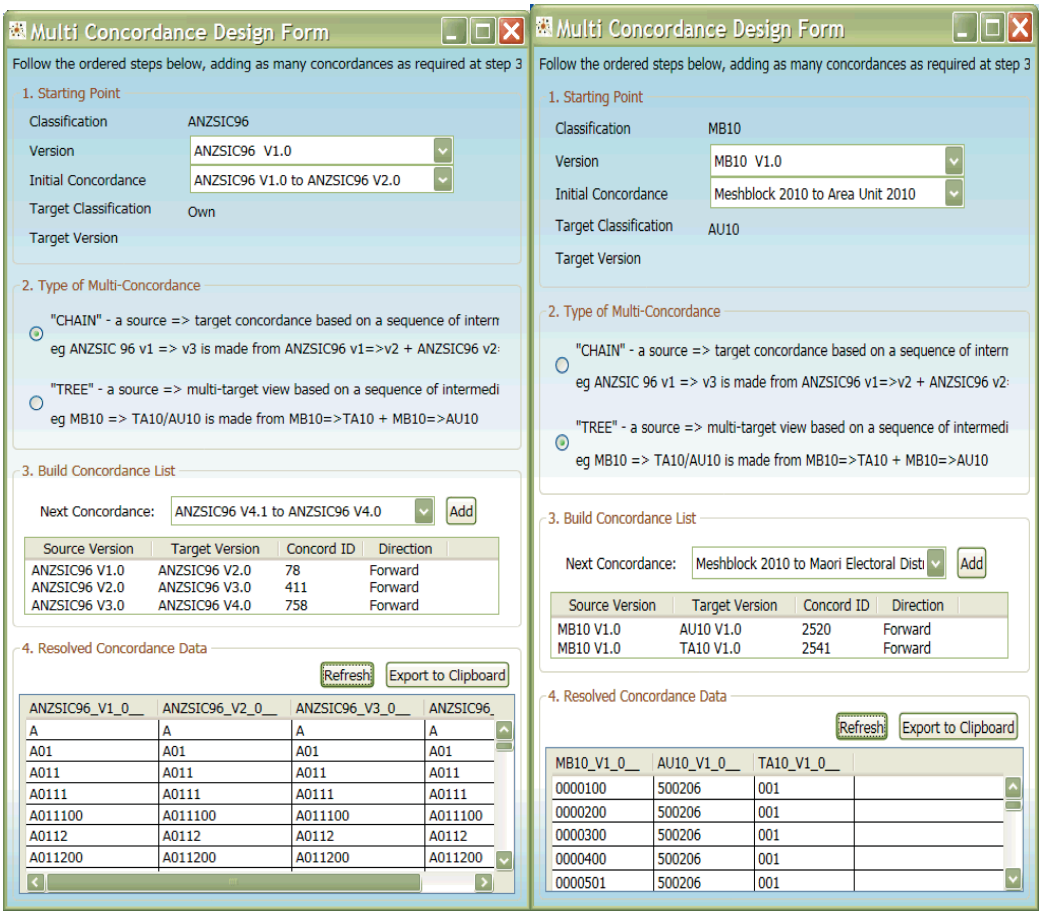
Multi-Concordances

One of the key functions that users have requested is the ability to create multi-concordances ie concordances between more than two classifications or versions. As stated earlier the current CARS system uses a Microsoft Access application to enable the creation of multi-concordances and this process has been built into the CARS PoC model. There are two types of multi-concordances available. These are:

- Multiple versions of the same classification
- Multiple versions of different classifications

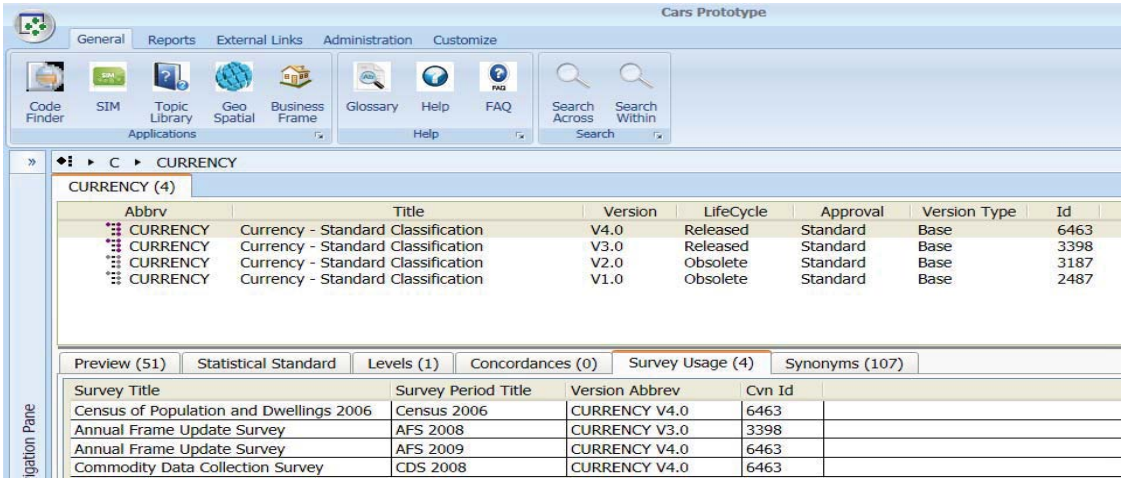
The former enables the creation of a complete time-series for all codes contained in a classification and this is particularly useful where the conceptual base and structure of the classification has not changed. The latter option enhances the capability to link different but related classifications, for example industry to product to trade. This may facilitate the identification of common areas across classifications that may lead to a rationalisation of existing standards.

A design wizard is used to create the concordances and this is illustrated below. Upon completion the multi-concordance can be exported into applications such as Microsoft Excel.



Survey Linkage

One of the more important features to be explored in the CARS PoC is survey linkage functionality. The benefit of this function is that it provides information on which surveys are using which classifications and facilitates the monitoring of the use and compliance of standard classifications.

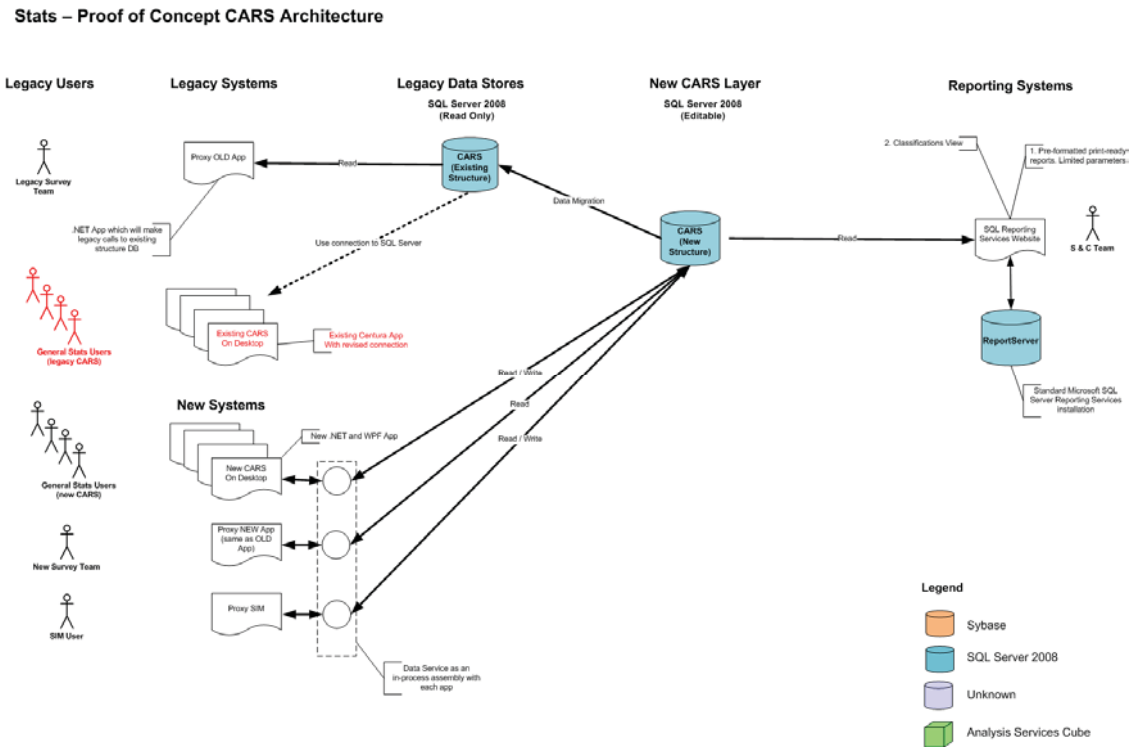


Additional Features

Other features required for future classification management are:

- the ability to notify internal users of classification changes or updates by automatically creating an email with the relevant contact persons populated in the 'To' box, based on the survey linkage information;
- better integrated linkages with internal IT systems, .NET survey systems and external websites
- storage of statistical standards, which contain the supporting metadata for a classification, and web enabled publication of these.
- Web accessibility for viewing and managing classifications

Appendix 1: Data Architecture Model of CARS Proof of Concept



Appendix 2: Possible Production Architecture for a new Classification Management System